DOCUMENT RESUME

ED 319 772 TM 014 977

AUTHOR Tucker, Susan A.; Dempsey, John V.
TITLE An Evaluation Semiotic.
PUB DATE Apr 90
NOTE 36p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Case Studies; *Evaluation Methods; International Programs; *Likert Scales; Models; Perception; *Program Evaluation; *Semiotics
IDENTIFIERS Stakeholder Evaluation

ABSTRACT

The semiotic interpretative context is extended to program evaluation. Assumptions about the semiotic process of constructing, describing, and judging reality are examined in terms of the viability of the process in explicating a perception-based evaluation model. Models from which inspiration was drawn include R. W. Tyler's instructional objectives model; M. Provus' discrepancy model; the Context, Input, Process, and Product model; M. Scriven's Goal-Free model; and R. E. Stake's model of evaluation. Five core evaluation signs are analyzed in terms of a recently completed case study of perception-based evaluation of a cross-national cultural awareness training program funded by an international foundation. Finally, a semiotic analysis for evaluation studies is discussed in terms of 13 Likert-scaled polemics. It is concluded that evaluation should be approached as both a native undertaking and an alien undertaking, and from an insider viewpoint and a disciplined outsider perspective. What is needed is a paradigm that negotiates major questions openly with all participants, gathers quantitative and qualitative scores of evidence, and applies standards or value criteria that are neither elitist from the perspective of the evaluator nor captured by the tacit agendas of the stakeholders. Two tables and a 72-item list of references are included. (TJH)

AN EVALUATION SEMIOTIC

Susan A. Tucker

and

John V. Dempsey

University of South Alabama

Paper presented at the 1990 Meeting of the American Educational Research Association, Boston, MA.

Running Head: An Evaluation Semiotic

Background

According to the thesis articulated by Graubard (1972), the task for educational reform and social change is not to initiate further studies but to equip individuals with the means to understand and struggle against structural and attitudinal impediments to implementing known solutions. Only an integrated and balanced division of effort along diverse psychological, social, political and economic fronts can enhance the value and utility of education and focus on basic priority needs of people. But the challenge still remains--how to transcend the rhetoric of "integrative" or interpretive approaches into action.

Sebeok (1986) offers semiotics as a context for reconceptualizing foundations in the human sciences given its ability to transcend dichotomous and reductionistic methods of perceiving with a world view that is holistic, respects complexity, and fosters synthesis. What the semiotic perspective can mean for educational researchers was recently analyzed by Driscoll and Flynn (1989). Specifically, they identified five assumptions of a research semiotic:

(1) systems are open and life is disorderly;
(2) behavior is modulated by immediate environment, species history, and individual history;
(3) the eventual effects of signs is what constitutes meaning, for the individual, a community and ultimately what is accepted as "truth". In fact, Cunningham (1989) advocates "the arbitrary nature of these signs is not readily apparent to the human organism until they are exposed to cultural systems which depart from their own";
(4) unities are assumed, not dichotomies (e.g., objective/subjective, qualitative/quantitative, and individual/system;
(5) reality is considered as the "final settled opinion of an indefinite community of scientific investigators," because we humans "can never get closer to that ultimate reality than our perception and reasoning powers can take us."

Purpose

The purpose of this paper is to extend this interpretative context to program evaluation. Assumptions about the semiotic process of constructing, describing, and judging reality (i.e., our Umwelt) will be examined in terms of its viability in explicating a perception-based evaluation model (Kunkel & Tucker, 1983). Five core "signs (e.g., the interpretive process inherent in major questions of focus, sources of evidence and standards of quality) will be analyzed in terms of a recently completed case study of a perception-based evaluation of a cross-national cultural awareness training program funded by an international foundation. And finally, a semiotic analysis for evaluation studies will be discussed in terms of 13 Likert scaled polemics.

Even with extensive developments in the fields of research and evaluation, most decisions about educational programs continue to be made without systematic evaluation. Annually, decisions about budgets, current programming, and future planning are often determined by

political expediency. Often these decisions are based on unfounded opinions and insufficient information. What occurs is a form of evaluation, but evaluation without a theoretical base, without systematic methodology, and where values and judgments are egocentric, elusive, and/or covert.

Since the late 1960's, authors in program evaluation have addressed the need for systematic models. Undoubtedly, the use of models to order educational processes facilitates the conception of problems and the perception of interrelationships within these problem contexts. While each model may view the same set of interrelationships, it is inevitable that certain concerns of the model builders will differ" due to personal frames of reference (Alkin, 1968). Dillon (1981) extends this personal dimension of modelling to include "prefabricated configurations of concepts" or " tterns of coherence that we expect to find".

The modeling of social systems involves another potential danger. Education's use of analog models drawn from the "hard" sciences is unsound to the extent that empirical models represent only "observable" realities. The human dimension in education requires a model which represents both observable and subjective realities. And the dimension of change inherent in evaluation requires a model which relates values, myths and paradigms.

As early as 1971, Daniel Stufflebeam raised several issues of concern to evaluation change agents--issues which have yet to be adequately addressed by prevailing evaluation models today. He notes four problems in the application of experimental designs to evaluation. First, the use of experimental designs conflicts with the principle that evaluation should facilitate continuous program improvement. According to a colleague of Stufflebeam, "experimental design prevents rather than promotes changes in the treatments because ... treatments cannot be altered if the data about differences between treatments are to be unequivocal" (Guba, 1969, p. 34). Secondly, traditional research methodologies are useless for making decisions during the planning and implementation of a project. Rather, the stress on controlling operational variables creates a contrived situation and blocks the collection of natural and dynamic information. Thirdly, the typical research design does not apply to the "septic" conditions of most evaluation contexts. That is, the evaluator is not interested in establishing highly controlled conditions within which universally true knowledge can be generalized. Instead, "one wishes to set up conditions of invited interference from all factors that might ever influence a learning (or whatever) transaction" (Guba, 1969, p. 33). As a final point, internal validity is gained at the expense of external validity. Clearly, equating evaluation models with empirical models is destructive to dynamic program development.

In essence, we are considering the semiotics of improvement-oriented evaluation models. Porsecanski (1989) captured this dilemma well when he said "individuals and groups do not really accept changes that they do not desire or consider necessary, and that tradition, beliefs, and social

status generally prevail over change". This dilemma is further compounded in evaluation when the evaluator threatens to question existing structures and commonly accepted myths.

More on the Literature of Educational Evaluation Models

The current practice of evaluation relies heavily on three models developed over twenty years ago. This section proposes to examine these models in terms of three tasks basic to any evaluation involving culturally diverse contexts, i.e., how they determine what is to be questioned or studied, how they determine and systematically collect relevant data, and how they determine operating values and judge program worth.

Model diversity emerges from the various emphases placed on each of these tasks and from the evaluator's relationship to the client system in regard to these tasks. For example, Pace's (1968) analysis of evaluation models indicates that when the total unit to be evaluated is large, complex, and of long duration, a different model is required -- one that considers a broad range of social and educational consequences. Focusing on the assessment of program objectives must be expanded to question other program dimensions (e.g., expected and unexpected consequences, cultural characteristics of the setting, the process of program delivery). Evidence gathered to answer these questions should be both quantitative and qualitative, including the systematic collection of personal perceptions.

To demonstrate the different emphases placed in various evaluation concerns, the following section analyses three prevailing approaches: the instructional objectives approach, the decision-making approach, and the values-based approach. Special attention is directed to the presence or absence of several criteria: a systematic methodology which utilizes both quantitative ("hard") and qualitative ("soft") data sources to portray a more holistic reality; helpfulness toward program improvement; the evaluator's openness to make values overt and to facilitate the planning of probable futures for a program.

Instructional Objectives Approaches

Tyler is known for an evaluation model which stresses measuring student progress by means of instructional objectives. Originated over fifty years ago, he conceives evaluation as the process of assessing the degree to which a program is achieving its predetermined objectives. The model is not concerned with judgments of the value of these objectives as much as the need to spell out the objectives behaviorally. These objectives then serve as the goals for teaching and the basis for testing achievement. Ideally, measurement is broad in scope and continual. In addition, Tyler proposes teaching and evaluating both cognitive and affective outcomes. While his definition

focuses evaluation on "results," Tyler's (1942) vision of the purposes of evaluation is broader than products:

a. provision of a periodic check on the effectiveness of the educational institution;
b. validation of the assumptions upon which the educational institution operates;
c. provision of information basic to effective guidance of individual students;
d. provision of "psychological security" to school staff, students, and parents;
e. provision of a sound basis for public relations;
f. help to both teachers and students to clarify their purposes and to see more concretely the directions in which they are moving.

Using Tyler's approach, an evaluation consists of five basic steps. First, educational objectives are formulated and classified according to the level of required specificity. Second, objectives are defined behaviorally. Third, situations are identified in which the behavior could be expected to occur. The fourth step involves the technical methodology for appraising student behaviors. Once collected, behavioral data are interpreted and then recycled for design improvements. That is, intended objectives can be revised. Such improvements lead to the modification of teaching which in turn leads to another evaluation and revision and so on.

Tyler's model has several merits. It provides valid, reliable, and objective data for an evaluation. It allows the evaluator to indicate attained and unattained objectives.

On the other hand, strict application of the Tyler Model creates difficulties. It ascertains student outcomes but ignores the contexts and processes that lead to these final outcomes. The statement of objectives in behavioral terms is a long and often tedious procedure which limits sources of evidence to purely quantifiable data. In addition, it is too easy to avoid questions about the worth of the objectives themselves, particularly if the evaluation is carried out after the objectives are formulated. When this happens, the usefulness of the evaluation tends to be limited to superficial reformulations of performance objectives. Finally, the question of comparing various programs using Tyler's model is explicitly avoided.

In response to these difficulties, more flexible "neo-Tylerian" models have emerged (Taba & Sawin, 1962; Cronbach, 1963; AAAS Commission on Science Education, 1965). Taba and Sawin propose an evaluation model that focuses on the collection of information in order to determine why some students failed to achieve stated objectives. Tyler's data sources have been expanded to include observations on teaching method, patterns of classroom interaction, physical facilities, and student motivations.

"Neo-Tylerian" ideas have contributed to the field's shift from a terminal focus to one which synthesizes both process and product elements. The emphasis of evaluation is on facilitating curriculum improvement. In this author's opinion, two major limitations persist. One, the evaluators do not incorporate their clients' values into the proposed evaluation. Secondly, futures planning is neglected.

Decision Making Approaches

A second group of evaluation theorists believes the evaluator's task is one of delineating information to be collected, planning a data collection methodology, and helping decision makers use the information collected. It is the responsibility of the decision maker, not the evaluator, to make judgments from this information. Four questions are basic to this approach:

a. What should the objective be?
b. How should the program be designed to meet these objectives?
c. Is the program design being carried out effectively and efficiently?
d. Was the program worth what was invested in it considering the products achieved? (Reasby, 1974, p. 23).

Major decision making models to be discussed are: the Discrepancy Model and the CIPP Model.

A. Discrepancy Model

The Discrepancy Model, developed by Provus (1969, 1971) emphasizes the discrepancy between the ideal standards in a program and the real performances. Provus assumes that there can be no evaluation without a standard. The explicitness of the standard generally determines the precision of the evaluation. It is the task of the evaluator to assist and train the program staff in evaluation. Evaluative training should emphasize continuous and cooperative involvement between the evaluator and his clients. Special attention is paid to training the staff in: 1) delineating and agreeing on program standards; 2) awareness of political and scientific considerations; and 3) determining whether a discrepancy exits and using this information in a feedback loop to identify program weaknesses.

According to Provus, education programs must be evaluated in stages relative to the development and stability of the program being investigated. Five stages of evaluation have been posited in the Discrepancy Model: design, installation, process, product, and cost. At each stage, comparisons are made and discrepancies sought between operational reality and some ideal standard derived from the values of the program staff and the client population it serves.

Provus (1971) concludes that the major weakness with the discrepancy model seems to stem from failure to understand the limits on institutional behavior set by field conditions, the intricacies of the decision-making process, and the origin and use of criteria in that process. Other criticisms include: the absence of a context evaluation and the notion that the use of behavioral standards may limit the creative, adaptive responsiveness of a program staff. Using Provus's approach is unclear who actively participates in program decision-making. Equally unclear is the type and ultimate use of information to be collected at each stage. The model could also be faulted for being incapable of evaluating rapid, large-scale changes. Its ability to evaluate more than one

program at a time is also questionable. Decision-makers are not always rational, yet the model assumes such behavior. Finally, the Discrepancy Model does not address how evaluators are recruited, trained, and supported in the system.

B. The CIPP Model

The Phi Delta Kappan Committee on Evaluation, chaired by Daniel Stufflebeam (1971), has perhaps exerted more influence than any group in attempting to lead the educational profession away from excessive reliance on classical research models. Instead, a decision-making evaluation model is offered which assists decision makers in pinpointing their values so that they can best be served by the decisions made. This model, known as CIPP, specifies four types of evaluation: Context, Input, Process, and Product (1983).

Context evaluation defines the relevant environment, describes the desired and actual conditions pertaining to that environment, identifies unmet needs and unused opportunities, and diagnoses the problems that prevent needs from being met and opportunities from being used. Context evaluation serves planning decisions in order to determine objectives. In order to "monitor" the total system, tools such as surveys, standardized tests, and demographic statistics are used.

Input evaluation provides information for determining how to utilize resources to meet program goals. It involves identifying and assessing three aspects: a) relevant capabilities of the responsible agency, b) alternate strategies for achieving program goals, and c) alternate designs for implementing a selected strategy. Data sampling methodologies might include committee discussions, research of the literature, and consultant assistance. Input evaluation serves structuring decisions by specifying a plan of action to achieve desired objectives.

Process evaluation provides periodic feedback to persons responsible for implementing plans and procedures. The strategy involves a) identifying and continuously monitoring the potential sources of failure in a project, b) projecting and servicing preprogrammed decisions, and c) describing what actually occurs during the program. Process evaluation serves implementing decisions in order that project operations can be controlled and the plan of action realized on a day-to-day basis.

Product evaluation's purpose is to measure and to interpret attainments. This occurs not only at the end of a program or project cycle, but as often as necessary during the program. In contrast to a terminal or "product" focus in Tylerian models of evaluation, the CIPP model considers product evaluation as a continuous activity since attainments occur throughout the program. Product evaluations specifically serve recycling decisions in order that project outcomes can be reacted to and judged. Furthermore, it serves the decision maker, who must decide whether to continue, to terminate, or to modify a program.

The procedure for the four evaluations mentioned above is basically the same. Firstly, identify the type or decision to be serviced. Secondly delineate the information needed. After the relevant decision makers have been identified, develop techniques to obtain and collect this desired information. The final step involves providing the information to relevant decision makers and audiences.

Though mechanically total, the CIPP model excludes overt values in its schema. According to Stufflebeam, the evaluator's role is one of collection, organization, and analysis of "relevant" data for "relevant" decision makers. The evaluator is not directly involved in program decision-making. Overt judgment of the intrinsic worth of the program's objectives is not considered by the evaluator or the clients. The entire question of values is kept tacitly in the decision-makers' domain. When the CIPP evaluator provides information to the decision-maker regarding which values will be served by alternative decisions but does not offer overt judgment himself, he is demonstrating the logical inconsistency of CIPP's theoretical stance on values. Though tacit, the selection of methods to uncover these values is largely a matter of the evaluator's values. Another limitation of the CIPP model is its inability to answer three basic questions. Firstly, how do evaluators and/or clients know which values are operant? Secondly, how can various effects of schooling be determined? Lastly, what processes are necessary to enable the decision maker to apply value criteria?

Values-Based Approaches

There are two major examples of values-based approaches to program evaluation. The first to be discussed is Scriven's Goal-Free Model. The second model to be discussed involves Stake's responsive schema.

A. Goal-Free Model of Evaluation

Scriven's (1967; 1972; 1978) Goal-Free model is in definite contrast to evaluators advocating objectives-based or decision making approaches. According to Scriven, the distinction between the roles and goals of evaluation are often intentionally blurred. Whatever its role, the goals of evaluation are always the same -- to estimate the merit, worth, or value of the thing being evaluated. Scriven goes on to point out that the subversion of goals to roles is very often a misguided attempt to allay the anxiety of those being evaluated. The consequence of this kind of distorted evaluation could be much more undesirable than the anxieties evoked.

Scriven divides evaluation into two areas, formative and summative. Formative evaluation is concerned with questions such as : What goals should the program achieve? What is the plan for achieving these goals? and Does the operating program, when developed and put into practice,

achieve the desired goals? Summative evaluation comes at the end of a program. Basically, it is an assessment of the finished product.

The distinction between summative and formative has implications for the personnel involved in the evaluation. The formative evaluator must work closely with the program manager. For the summative evaluator, quite the opposite is true. He must be free of any potential conflict of interest so that his evaluation has an integrity of design and conclusion.

Scriven (1972) declared that both summative and formative evaluations will increasingly become "goal-free." Believing intermediate goal evaluations to be unnecessary, Scriven stresses the evaluation of actual effects against a profile of demonstrated needs. One of the many roles of the evaluator is to examine the goals of the educational program and judge the worth or value of these goals against some standard of merit. From the data he has collected, the evaluator can determine whether these objectives are being met. Formative value judgments of the program are made for producers and summative value judgments are made for the consumer.

On many occasions, the staff evaluator is likely to have "tunnel-vision which respect to the effects of the materials (or methods, etc). That is, a tendency to look mainly in the direction of the announced goals" (Scriven, 1972). Therefore, it is essential that there is an external evaluator who has not been informed about the project's goals so that he may devote his unbiased attention to looking for actual effects or side-effects of a project. Thus, he may notice something that everyone else has overlooked or he may, produce a "novel overall perspective".

Several questions remain unanswered about "goal-free" evaluation. One major concern is how best to insure that external evaluators properly judge actual effects of a program, whether planned or not. What standards are there to judge whether a goal-free evaluator is not arbitrary, inept, or unscrupulous in his actions? How does one judge how well a goal-free evaluator has interpreted the "demonstrated needs" of a project? It seems these questions will not be answered until Scriven sets up the necessary dynamics of mutual negotiation between the evaluator and the evaluated. Only in this way can he realize a holistic evaluation schema that includes both internal and external values.

B. Stake's Model of Evaluation

Stake would undoubtedly agree with Scriven's feelings about goal-free evaluation. The theme throughout much of Stake's writings (1967, 1970, 1972, 1973, 1975, 1982) is that an evaluator must do his best to reflect the nature of the program and not focus on what is most easily measured. Furthermore, he writes less about the link between evaluation and decision making than other evaluation theoreticians.

Stake agrees with Scriven that both descriptions and judgments are essential and basic acts of any evaluation and should be combined to portray an educational program. For Stake, the role

of the evaluator is much more than an interpreter of test scores. He believes that the evaluator's task is one of identifying outcomes or "intents" that are contingent upon particular antecedent conditions and instructional transactions which occur between teachers and students, students and teachers, students and students, etc. These contingencies seek to discover the connecting relationships between transactions on two levels: first, what was intended; second, what is observed. In addition to contingencies, the evaluator examines the congruency of these intents and observations in terms of antecedents, transactions, and outcomes. Thus, descriptions are obtained through a process of presenting a rationale for the evaluation based upon the local situation, followed by articulation of intents and observations.

Stake also believes that the role of the evaluator is one of making subjective decisions in selecting variables for the evaluation. These judgments are reached by having "experts" generate standards for antecedents, transactions, and outcomes. Then, intents and actual observations are compared with these prescribed standards of excellence. Finally, the evaluator makes judgments of worth based upon these standards. Thus, the evaluator should be a specialist in the collection, processing, and interpretation of data. Moreover, she should be an expert in the presentation of data, to the client system. She should be familiar with other programs having similar objectives or intents, and she should provide the client- system with these comparisons. Stake stresses that data be collected from a wide variety of sources and that the evaluator's role is to assist the educational enterprise in a realistic formulation of objectives and in the assessment of the congruence of these with the observable data collected. The evaluator is a technician who can provide both relative and absolute judgment. In this respect, Stake takes a more conservative stance than Scriven. Stake (1973) recommends that the evaluator should not take an absolutist position regarding the program's goals since this is likely to make clients less willing to cooperate in the evaluation.

The Stake model exceeds Tyler's model in attempting to describe and to judge the entire educational enterprise, rather than examining outcomes alone. In terms of process and scope, the Stake model deals more comprehensively with a number of evaluation needs. The assessment of evaluation procedures in the descriptive matrix is unclear as it appears that procedures would be described with transactions. However, procedures are selected prior to the learning experience. An effective method for analyzing the role of values in the program can be found in this model as can a methodology for discovering which values are being served.

Like Tyler, Stake does not provide for evaluating decision alternatives during the structuring of the learning experience. Neither does he provide for adequate feedback in program development. While the educational program is looked at in terms of antecedents, transactions, and outcomes, underlying the whole approach is a perspective that looks at evaluation from the terminus. It does not forcefully facilitate dynamic development of educational programs.

The Need for a Theory of Evaluation With Explicit Quality Criteria

It appears current evaluation models can create methodologically sound strategies but credibility tends to be lacking unless some process exists for making the operational values of both the evaluated and evaluators overt. Furthermore, these overt values need to be incorporated into the evaluation itself. Support for this position has been persuasively advanced by Weckworth:

> "First, there is no one way to do evaluation; second, there is no generic logical structure which will assure a unique "right method of choice." Third, evaluation ultimately becomes judgment and will remain so, so long as there is no ultimate criterion for monotonic ordering of priorities; and fourth, the critical element in evaluation is simply: who has the right, i.e., the power, the influence, or the authority to decide." (1969, p. 48).

It is the purpose of this section to offer a theoretical model which strives for credible synthesis--which is especially difficult to achieve in contexts reflecting diverse client and audience values. It assumes that there are no theories -- only personal perceptions of theories (Polanyi, 1966) and this is what will keep the model dynamic . A model is therefore perceived as a body of interrelated criteria, which in the case of the proposed model are quality criteria. These (quality) criteria present view of a topic. The topic's focus is a function of the builder's personal frame of reference. Specifically, perceptions and judgments are a function of the author's Umwelt. By making overt the vulnerable judgments of both evaluators and those being evaluated, a synthetic picture of what the evaluation intends to focus on can be negotiated. Through negotiation, operative values can be embodied in the major evaluation questions asked. In addition, these values can serve as judgment criteria for decision-making.

In the perception-based model of evaluation proposed, certain value criteria are not negotiated in the sense that along with accepting the evaluator personally, the primary audience must accept five quality criteria inherent in the model. That is, the model embodies the following criteria: holism, helpfulness toward program improvement, evaluator vulnerability with the audience, acceptance of both "hard" and "soft" data sources to portray reality, and facilitation of planning a program's future. Additional quality criteria are negotiated with individual audiences.

Five Evaluation Signs

Analysis of perception-based models of program evaluation suggest five core "signs" which advocates perceive as essential to conducting and interpreting studies. In fact, the first four assumptions of Driscoll and Flynn's research semiotic are very congruent with the first four perception-based signs delineated below.

First, a holistic or comprehensive base is advocated rather than the reductionistic principles frequently associated with western dominated program

evaluation designs. Education and training are considered broader, socio-economic phenomena which require a synthesis of multiple perspectives, including psychological, social, political, anthropological, philosophical, and economic. Further, humans are regarded as creative, historical forces who are in need of a critical, reflexive way of knowing (Allo, 1961; Freire, 1970 a,b). Man is real and active to the extent that he *perceives* and *produces*. Active men are timebound and are conditioned by a determinate social structure and nature.

Theory should be *holistic* rather distort the total evaluation picture with an undue emphasis on quantification. That is, it goes beyond mere correlational analysis or statistical inference with the addition of an existential phenomenological/philosophical treatment wherein all syntheses are repeatedly tested against experiential perceptions.

There is no way of surveying the whole of reality since one has no assurance that he has experienced anything more than an infinitesimal sample of it. As different methodologies increasingly seek to make reality determinant, they become caught in their own closed system. A labyrinth of theories is constructed in an effort to reduce the excluded aspects to the included ones, but there is no opening in this finite maze. For example, the natural sciences have limited the conception of nature to that which can be empirically investigated, i.e., the mere description of observable experience. We do not object to the scientific method, but only to the regarding it as exhaustive of cognitive method.

Support for this particular criterion is especially strong in the work of French change theorist, Georges Allo (1961). Science, claims Allo, necessarily fragments reality when it abstracts. He challenges this procedure in the specific case of value studies because by definition values involve evaluational totality, not merely valuational fragments. A normative sequence which delineates existing value constellations and explores alternatives in terms of probable future is a result of Allo's examination of third world contexts. One can see parallels in Allo's research methodology to the pedagogy of the oppressed devised by Paulo Freire (1970a). Both have directed their efforts toward the reduction of ethnocentrism and attempted a holistic orientation by means of testing knowledge through action which is critical, dialogical, and reflective.

Second, while promoting the value of commitment toward program improvement versus proofs inherent in objectives-based evaluation models, it is recognized that practitioners of this approach can sometimes experience difficulty avoiding alienation of diverse constituencies (e.g., funders, practitioners, informants). Evaluators must also be eclectic and capable of listening to divergent values so that a truly "synthetic" base of program improvement can be perceived as helpful rather than imposed. *Negotiation* among all the key constituents (i.e., trying to achieve 360 degrees of representation) is advocated throughout the evaluation process, including negotiating the initial evaluation paradigm and its major questions of focus.

The negotiation should be helpful in terms of facilitating the commitment of all parties concerned to program improvement. The values of the evaluator and the evaluated are exposed by overt, dynamic dialogue, negotiation and pedagogical action. By placing responsibility for the change effort directly in the hands of the evaluated, it becomes increasingly probable that these change pedagogies will enable the understanding of a system's relative value positions in their totality. Reconstruction or improvement of the outer boundaries of a "society's existence rationality" can then result while its "core values" remain respected and intact. Overall, this negotiation process becomes especially challenging when the evaluator's "core values" differ from his clients, sources or consumers.

Third, avoid a reductionistic "research" perspective due to the inherent danger of discrediting obvious localized learning and grass roots data in favor of "hard numbers". Normative or criterion-referenced approaches dominate and there is no question that trying more synthetic models is a risky business--politically, economically and theoretically. If we assume this risk, creative divergence gained from a unification of qualitative and quantitative sources of evidence (as opposed to data collection strategies of control theories such as Tyler's) becomes an actualizing component of potential realities rather than the status quo.

The model can then *duel with the positivist and reductionist biases* inherent in empiricism by granting primacy to the realm of synthesis and the totality of experience. The limitation of analysis can transcend a theoretical framework of deductive hypotheses and quantification as the only test of truth. Abductive and inductive frameworks can be included in the model. As Goulet (1971) has cautioned: "It is true that quantitative changes of a sufficient magnitude may produce changes in quality, but even in these cases, the uniquely qualitative dimensions of reality tend to escape the purview of those best equipped to measure or describe quantitative facets of a whole. Consequently, one may reasonably doubt the ability of social scientists to perceive qualitative manifestations of value change except through the distorting lens of their own positivist bias (p. 219). Thus, for our purposes, abstraction and analysis are applied only at certain moments in a specified sequence.

Fourth, the concept of evaluator vulnerability is perceived positively rather than as something to be defensively covered up. Vulnerability is used in the sense of actively seeking multiple perceptions of key actors, whether in context, process or product stages in order to achieve the very unsteady state of intersubjectivity. Evaluators are seen as *objective* rather than neutral (neutrality being equated with stasis or death). Basically, it appears vulnerability can be characteristic of an effective change agent, but the premise does not extend as far as Machiavelli's end of the continuum with his contention that this openness is a necessary illusion. Pharies' (1985) interpretation of semiotic reality as "the final settled opinion of an

indefinite community of scientific investigators" would be eschewed as elitist and a more open systems view of shareholders included in the process of defining and judging a program's reality.

The theory should deal with the *elitism of the evaluator* as the possessor of distinct skills and a quasi-mystical database in order to avoid manipulation of the evaluated. In his exploration of the assumptions of Western social scientists, Goulet has discussed the damaging ramifications of elitism. He contends that "researchers practice a new form of clericalism -- not of ordained clerks, but of methodologists initiated to the exclusive task of explaining reality's crucial dimensions. No matter how insistently an individual researcher disclaims elitist views, he is in fact the possessor of a mysterious tool-kit, the contents of which cannot be fathomed by the 'objects' of his study. The stance of the researcher manifests something of the 'magical' posture described by Levi-Strauss (1969, 1976, 1979) when he speaks of the power which accrues to anthropologists investigating primitives who lack mastery over the written word.

By recognizing that one's perceptions confirm, continue, enlarge, and eventually, correct one another, it is possible for the evaluator to synthetically enter into the perceptual systems of the evaluated. Disagreements may well occur. These are removed, at least they are considered as removable by discussion and mutual criticism. According to Edmund Husserl (1950) the consciousness of a common perceptual world as the universal horizon within which all of us live arises through intercommunication. "I may and do perceive the same thing as my fellow man, but the same thing presents itself to each of us from a different side, under a different aspect and perspective. Every such perceptual presentation implies in its inner horizon references to further perceptual presentations, all belonging to one total system" (Gurwitsch, 1966, p. 430). By being open to the multiplicity of perceptual experiences of both his own and those being evaluated, the evaluator is not trapped in a narrow kind of consciousness of egocentrically determined utilitarian material. Recognition of the potential of intercommunication in contributing to an accurate picture of reality has been articulated by Aldous Huxley (1963) in The Doors of Perception.

> "Every individual is at once the beneficiary and the victim of linguistic tradition into which he has been born -- the beneficiary inasmuch as language gives access to the accumulated records of other people's experience, the victim insofar as it confirms him in the belief that reduced awareness is the only awareness and as it bedevils his sense of reality, so that he is all too apt to take his concepts for data, his words for actual things. That which in the language of religion, is called 'this world' is the universe of reduced awareness, expressed, and, as it were, petrified by language. The various 'other worlds,' with which human beings erratically make contact are so many elements in the totality of the awareness belonging to Mind at Large" (p. 23).

To paraphrase Merleau-Ponty (1962), language is *verflochten* with our horizon upon the world and humanity. Language is borne by our relation to the world and to others which in turn supports and creates it. It is through language that our horizon becomes open and 'endlos.' Thus, only by developing the evaluation in a dialogical way so that both the evaluated and evaluators participate

actively in overtly defining the ground rules of program improvement can manipulation be prevented and an accurate picture of reality articulated. Only through commitment to seeking multiple realities of truth can the evaluator be a responsible decision-maker and future planner.

Finally, a positive view of futures as part of an evaluator's Weltanschauung is maintained. Sebeok's notion of behavior being modulated by one's immediate and past environments is recognized, but perception-based evaluation consciously develops a sense of futurism as an essential truth. In essence, we are referring to Peirce's notion of a sign of possibility which is represented by its interpretant. Unfortunately, present "realities" are often used as an excuse for inaction. One step in pulling evaluators from this abyss could include two interpretive strategies culled by Driscoll and Flynn's guidelines for conducting educational research: presenting evidence in ever expanding networks of implications (Campbell, 1984) and making only modest pronouncements given truth is not fixed and making judgments of usefulness or effectiveness within interpretive contexts (Cunningham, 1987).

The topic of probability offers a further polemic when one considers the significance of judgment as a "willed reality," One can argue that empirical research is based on a probability of truth, while its counter, evaluation, values individual perceptions as central sources of data and judgment. Probability plays a key role in both evaluation and educational research by developing a game of signs. Both evaluation and research follow certain rules and are contained within certain boundaries which create a sense of order. Players in each game like to delude themselves with the belief they are operating in a realm independently of extraneous variables, i.e., independently of "ordinary" life. The players flaunt certain credentials as possessors of an exclusive information base. Educational researchers more blatantly sacrifice meaningful data to reductionistic and positivistic analysis. Program evaluators who adopt the researcher game plan are more subtle and ultimately more distorting of the "ordinary" reality they attempt to capture and judge. Evaluators must be careful not to confuse the characteristics of a good evaluator with a good magician. Drawing upon the work of the magician Blackstone, Baron (1987) lists seven rules for good magic which program evaluators have been know to abuse on occasion:

- use conversational skills to achieve misdirection
- play on audience's presuppositions
- play on audience's expectations
- never tell an audience how a trick is done
- the best tricks are those in which the audience has an active role.
- the effect is what matters, not the means of achieving it.

An empirical sign only partially reveals its dynamic object and the extent to which the dynamic object is revealed is the immediate object. The dynamic object is what actually determines the sign. It is the essence of the program we seek to judge. How does the evaluator see beyond the immediate object to the dynamic object? It is the dynamic object that has determined the

descriptions but the revealed descriptions are not sufficient in themselves for judging the dynamic object. The evaluator must depend on "collateral experience to grasp the true significance of the sign and make the correct diagnosis" or judgment (Houser, 1987, p. 265).

Another distinction between these evaluation and research signs can be seen in empiricism's tendency to develop as a closed system directed toward refinement of previously stated objectives versus holistic program evaluation's proactive orientation to the novel and open future. To the empiricists, research outside of the positivistic paradigm is valued only after a full-blown crisis destroys any viability for the established principles of the scientific method (Kuhn, 1962). Currently the reactive nature of empiricism is shaking the structure of science, with the resulting realization that future planning is important and "possible in fields lying outside of the bounds of absolute scientific causality . . ." (Gordon, 1969, p. 73). Thus, a proactive evaluator must base his conclusions and recommended course of action on his judgment of an improved, *better* willed reality.

An Evaluation Semiotic: Perceptions and Polemics

Evaluation involves the interaction of signs, objects and interpretants. Evaluation questions are the tools that enable us to present unfamiliar, change-oriented descriptions and judgments (dynamic objects) in terms of familiar, accepted descriptions and judgments (immediate objects). Descriptions and judgements are grounded in perception. That is, perception is considered the central source of evidence which permits the interpretation of data in terms of systematically ordered concepts. In a *specific* sense, perception is defined as what individuals report as their descriptions of events, activities, and values. Quantitative definitions of perceptions such as one finds in neurological and sense datum expositions have been found to be limited for our purposes, and the necessity of a definition incorporating Polanyi's (1966) 'tacit dimension' has become increasingly apparent. That is, "we have seen how the quality of the senses are both determinate and indeterminate, and, therefore, how a reality is directly present to consciousness, but present only partially. Such is the character of the world apprehended by subjectivity." (Earle, 1955, p. 99). Bergson uses the metaphor of a perceptual chain: "we can . . . conceive of succession without distinction and think of it as a mutual penetration, an interconnection and organization of the elements, each one of which represents the whole and cannot be distinguished or isolated from it except by abstract thought" (1910, p. 101). Schutz (1967) has expanded Bergson's concept of consciousness as an on-rushing stream of prereflective experience, which man lives through rather than knows about, and which is tied to some concept of time. The similarity here with Merleau-Ponty's imperative to philosophy to return to perception, that is to prereflective and prescientific experience is considerable. And the similarity increases when Schutz's accounts of the understanding of other people, and the social world in general are considered. And just as the

analysis of perception is, for Merleau-Ponty, analysis of the world-as-perceived, similarly for Schutz, analysis of subjective social knowledge is an analysis of the social world-as-known. For both phenomenologists the reference of philosophical analysis to the subject involves necessarily a descriptive and analytical recapture and reconstruction of the objective world, the objective correlate of subjectivity.

Moving theories into action results in paradigms. As Kuhn said (1970), the scientific paradigm is like a common door through which scientists must pass in order to understand their world and design new research. But most of these researchers are not conscious of the reductionistic assumptions imbedded in their paradigm, nor are many evaluators. Evaluation paradigms or myths consist of dynamic interactions between predictive and polemical relations. Signs in evaluation must take into account questions (contextual, process, product), sources of evidence, and qualities/standards by which to judge the resultant data and questions. These connections between questions, sources and qualities are not explained by real objects but rather *perceptions* of connections and objects. According to Riffaterre's (1988) notion of hypersigns, this connective process occurs through extension or deduction from a prior sign, which has the effect of imposing a re-reading and meta-mapping of that sign. For example, when we look at the connection between evaluation contexts and evaluation products, we need to examine both *anticipated* or ordinary signs and *unexpected* or extraordinary signs.

At this point it should not be forgotten that a person does not report a 'purely objective' description, only good or bad, better or worse descriptions which involve selective and systematic distortions. An evaluator must consciously deal with the fact that individuals selectively perceive, interpret, and place meaning upon their world, and then act accordingly. In fact, the solid facts of society are these perceptions, definitions, interpretations, and meanings which enable the perceiver to operate effectively most of the time in the face of incomplete informational input or under conditions of impaired observations. Franz From (1971), a Danish psychologist, believes perceptions of another's actions depend upon our familiarity with such factors as the type of action sequence played, situational norms, the actor or persons of his type, and whether we know something about how this kind of man usually behaves. As our familiarity with these factors becomes increasingly vague, "we see that others do something or other, but not what they are doing, or the experience becomes very disjointed without any connection or transition between the details. There may in such a case be given a more comprehensive, rather imprecise, sense." (p. 91).

The intention, even beyond our explicit thoughts, has been to open the sphere of reality to whatever we know and therefore, beyond the constrictions of quantifiable sense perceptions of physical substance. This view permits as 'real' the concepts of subjectivities, selves, and

personalities. All have a mode of being distinctly peculiar to themselves, and reducible neither to physical existence, biological life, sheer duration, nor to instantaneous actuality.

In a general sense, perception refers to what someone judges in terms of the overall theoretical construct. Judgment is viewed as the subjective act which wills something as true, in effect, a "willed reality" (Earle, 1955, p. 130) which usually occurs in the sphere of the indeterminate where probabilities concerning the truth of realities operate. Judgments are considered real only to the extent that a person's reports truly reflect his perceptions. It does not have the truth-revealing character of apprehension. Judgment thus is a very secondary sort of thing, completely dependent upon perception for its confirmation. Problems in program decision-making occur when something, someone, etc., is incapable of being directly perceived, in which case only a vague report of the operating perceptions can be gathered. Then, judgments about their particular qualities and their mode of being must be grounded upon not a single report, but upon a number of perceptions-descriptions from a number of subjects.

In considering the need for judgments in program evaluation, a more extensive examination of judgment is necessary. The concept of judgments as a "willed reality" evolves from imagination and personal valuing. Drawing from the actual world and provided by experience, imagination is a way of discovering and expanding existing potentialities, i.e., to really stand under the visible and direct and obtain meaning. The extent that possibilities revealed by imagination can go beyond the superficial and trivial, through the dynamic and creative powers of the evaluator, determines the extent of utilizing the deepest character of the program evaluated and expanding the arena of the new "willed reality," or vision of the better; the greater the range of major questions asked, the more legitimate the evaluator's judgments in making the third box and its willed realities the possibilities of others. This component, therefore, becomes an actualizing component of potential rather than a component of the description of status quo. As Polanyi has stated ". . . to see a problem and undertake its pursuit is to see a range of potentialities, believed to be accessible." (1966, p. 37).

What seems to be operating here are levels of potentialities determined by degrees of imagination. Indecision and lack of active change seem to be based on man placing on himself all kinds of limitations to action, analogous to the instincts of animals that determine the behavior sequence in advance. Specifically, that which is in accordance with the rules is good and acceptable because it gives peace of mind, while actions which contradict the laws are bad and undesirable because they are unpredictable and create uncertainty. This unrealistic orientation accords primacy to as objective and empirical a world as can be controlled, and bans the place of words like *subjectivity*, *imagination*, *tacit knowing*, and *intersubjectivity*. In reaction to this behavioristic tendency, the perception-based evaluation semiotic as presented in this paper has endeavored to synthesize the overt judgments of those operating as evaluators and the tacit value stance of those

being evaluated. Furthermore, through a formative negotiation process, predetermined operative values can be embodied in the major questions asked and evidence sought. Finally, an improved "willed reality" can be planned.

Paradigm of the Perception-Based Evaluation

The total unfolding nature of perception-based theory can be seen in the three column paradigm in Table 1. The paradigm calls for identification and negotiation of major questions to be asked (see Column 1). This might include both those cited or intended and accidental objectives. The second column of the paradigm attempts to describe an "accurate picture" of the program, ideally allowing reality to be described holistically by others and mutually agreeing upon these descriptions. The third column articulates the values and the rationale for making judgments. In fact, inherent in all three columns are five signs: holism, helpfulness, evaluator vulnerability, acceptance of both quantitative and qualitative dimensions of reality and a vision of the future. Negotiated judgments arising from the evaluator's *and* participants' egocentric value orientations must be made overt in the process of determining major questions. In describing an accurate picture of reality, some judgments in terms of validity and reliability are involved. The third column synthesizes these elements by providing a meaningful perception orientation upon which holistic judgments about programs are based.

Stepping from opinion and rhetoric to systematic evaluative inquiry forces a duelling with the description versus judgment polemic. Assuming all thought is in signs, what operates is a multi-tiered sign system dependent upon interpreting the descriptions of all the critical stakeholders in the evaluation. Sign fluency appears to involve balancing two dimensions of perception: the active (i.e., low filter, holistic, tacit, iconic) and the passive (i.e., high filter, part, explicit, verbal).

To make holistic judgments based on the facile reading of this multi-tiered sign system requires the initial step of negotiating major questions to focus the evaluation inquiry with stakeholders; "interaction foregrounds similarity" (Ricoeur, 1985).

Different than non-negotiated research questions, negotiated questions serve as indicants and metaphors. "Indication...seems to be a way the mind operates in complex states of affairs. Through indication we communicate what we cannot conceptualize. As a mechanism for indication, metaphor expresses intuition--mental consequences of perception which one cannot formulate with the words and syntactic constructions of one's language. The sense of the whole is simple. If "the poet can find metaphors that render these obscure parts accessible and eventually understandable" (Benzon & Hays, 1987, p. 60), then the evaluator must acquire similar skills to render the obscure accessible and be able to perceive both the focal and residual elements of a question. Metaphor leads to recognition and training in reading signs in first one's own culture

and then simultaneous blendings of multicultural metaphors should lead to a heightened awareness in other cultures. Specifically, Driscoll and Dempsey's (1988) application of the rational set generator may be one way to explicate cultural metaphors at increasingly complex levels.

How can the same sign (of possibility, fact or reason) mean so much more to one evaluator than another? Houser (1987) accounts for different interpretants based on the immediate objects (e.g., sources of evidence) being different. Understanding the immediate object depends on previous interpretants of previous signs--that is, background knowledge. Making judgments about the data depends on predefined standards or qualities and background knowledge. If our immediate objects are primarily quantitative, our judgments will convey a decidedly empirical flavor. Empirical conveyance is an analytical process whereby clarity is dissected and a scientific paradigm dismantled into ontological and methodological assumptions.

More holistic evaluation paradigms must examine what standards or qualities of "good" evaluation are. Prior efforts to predefine "good" evaluation processes are signified by the following words, many of which encompass but are not limited to reductionistic evalution models:

broad-based	integrative	multi-disciplinary	comprehensive
reinforcing	choice	personal	open-ended
analytical	developmental	challenging	promotes self-understanding
critical	dialogical	systematic	affective & cognitive
objective	subjective	intersubjective	ongoing
process-oriented	product-oriented	planned	futuristic
negotiated	diverse audiences	multiple clients	accessible
site-specific	realistic	improvement	problem-solving

By guaranteeing intersubjectivity, (which incorporates objectivity and subjectivity) this paradigm can actualize itself in the language of individuals' personal and holistic descriptions, observations, perception grounded decisions as to what to inquire into, and recommendations for modifications. This language recognizes both the limiting and expanding nature of man's conscious interactions. At a prereflective level of consciousness, man can be observed as living-through an onrushing stream of undifferentiated experience rather than *knowing about reality*

---

insert Table 1 about here

(Husserl, 1964; Merleau-Ponty, 1965). However, while the subject cannot grasp his own immediate flow of experience in the present, the paradigm accepts Schutz's (1967) argument that another self (possibly the evaluator) can meaningfully catch the other self's present flow of experience. In Bergson's terms (1910) there is a simultaneity between my stream of consciousness, my lived experience, and yours. I experience your spontaneous experience directly as you talk, for instance, but you can only capture your own spontaneous after it has flowed away in a self-conscious reflective turn.

The import of Bergson's and Schutz's ideas lies in the often neglected communication concept that our knowledge of others can be more direct than our knowledge of ourselves. The implications of this concept of intersubjectivity on paradigm can be seen in the expanded potential of meaningful reality assessment, both by the evaluator and the evaluated. Only by discovering and making evident through language both the evaluated and evaluator's presuppositions about valuing and meaning can the evaluation proceed. Without going into this subject further, it can be seen that Garfinkel's (1970) concepts of "indexicality" (i.e., all language and meaning depends upon conversational context and upon the interpretive work speakers and hearers do in conversing) and "reflexivity" (i.e., it is the ongoing attempt in conversational contexts to specify the meaning of those contexts, to formulate what the conversation is about and what is its purpose) speak to the dialectical relationship between man and language, between consciousness-purposiveness, and its linguistic form and content.

The evaluation case study

The following evaluation case study was selected because it exemplifies perception-based tenets. It also has the added feature of including contexts where key actors (e.g., funding agency staff, trainers, trainees, most other key informants and the evaluator) are exposed to cultural systems which depart from their own. Increasingly, semiotic experts have supported cultural interaction as a significant mediator in sensitizing individuals to perceive signs.

Since 1982, this bi-national foundation has funded nine projects for an average of three years each in order to:

- build a regional network of precollegiate educators trained to incorporate US-Japan relations into the curriculum;
- develop and make quality instructional materials available with each region;
- design model training and materials development centers operated by state-based teams.

The foundation contracted with the first author for its external evaluation. Conducted over a seven month period (May through December, 1988), major questions as well as standards of quality were negotiated between the evaluator and the funder. These major questions addressed

three phases of the teacher training program: context, process and product. Specifically, six questions were asked generically as well as other site-specific questions:

1. What are the purposes and unique characteristics of each center?
2. How have training and materials been developed overtime and across centers?
3. What are factors are contributing to ongoing success and failure?
4. What changes are needed in the overall program?
5. How effective have the centers been?
6. What policy issues have emerged?

Diverse sources of evidence were used to describe each question. Data sources included: review of project proposals; foundation project officer files and curricular/training materials developed by each center; analysis of internal evaluation data developed by center personnel; visitation to 24 sites throughout the US regions and four study groups in Japan; individual and group interviewing with 154 teachers, 128 students, 20 staff, 36 embassy and business leaders, seven advisory board members, 11 staff in other foundations, and four foundation staff.

The evaluator must analyze parallel and convergent signs, collect descriptions of the same event in American and Japanese semiotic systems, and make judgments regarding the paradox of simultaneous decoding of discrepancies and incompatible meanings. The speed of cultural change, the impact of technology and the inescapable lure of comparison between the two semiotic systems makes sharp analysis difficult. Both societies have witnessed tremendous changes, overt and subtle, since World War II. For example, what standards do we use to frame judgments about the extent of cultural awareness training, a goal all the funded centers have in common? For our part, we will seek to apply *core values* according to Allo's (1961) suggestions.

A good place to start the analysis is to examine several foundation sponsored summer study tours for overt and covert signs of cultural awareness. Imbedded in the travel plans are evidence of the degree of "cultural awareness" of its developer as well as the presence of stereotypical experiences versus deliberate tests of tourist rituals that challenge cultural assumptions.

To approach cultural awareness, one must be able to look at discrepancies of events between two cultures. Initially, this is at an implicit level. Benzon and Hays (1987, p. 64) talk about "the initial discrepancy between basic global physiological schema leads you to consult a representation which is built upon physiological parts and parts of parts, one which is a propositional construction of these physiological parts. You then search for parts of parts where the discrepancy is the greatest, that will tell you what has changed" An example of this is the discrepancy between pre-trip images of a country and post-trip images.

The second phase of developing "cultural awareness" demands an explicit awareness of assumptions embodied in one's own culture and another. What about those situations which are not aware of a discrepancy because you have been so captured in your own culture? We posit that awareness appears to be partially explained by Piaget's notion of that the readiness to perceive and

create inner representations of objects and events drawn from the external world is developmentally structured.

Moving from pre-trip assumptions and initial site dialogue to being able to explain discrete factual assertions entails both an *aesthetics of communication* and *objectification of discourse.* (Geertz, 1973). In the course of relating cultural judgments and cultural descriptions, the study tour member is confounded by his/her undefined role. Not quite a tourist and not quite "there on business", the tour member is particularly vulnerable to *discursive sabotage* (Jules-Rosette, 1989) where the accuracy of cultural awareness relies upon the content gleaned from the participant's inquiries. The aesthetic aspect of this process relates to the style in which information is communicated first by the native source to the study tour member and then by the tour member/teacher to a broader audience of students back home.

Harvey Sacks (1972) outlines rules of economy, adequate reference, and consistency to explain this discursive process. Through an adequate reference, a term appears as an appropriate and recognizable member of a category. Under the consistency rule, the term must be used in an uniform manner to classify members of the same set or collection. Consider the Japanese category, "sensei", or the American category, "Southerners" as a collective identification. You can go on and inventory a number of cultural traits according to which these people may be classified. These traits are classed with reference to livelihood, material culture, social arrangements, and political institutions. The danger of traits is that they can be easily translated into generalizations to support the tour member's decisions for assigning a particular identification to a category of people, objects or events.

Understanding another requires reciprocity and playing common game rules. This game involves playing upon a set of mutual correlations, or semiotic oppositions, in a manner that hinders attempts to transform dialogue into factual data about lineages, kinship, and village social structure. Though both are labelled *First World* countries, the East-West world view continua makes it particularly difficult for short-term visitors to transcend more than the superficial rules of each host country. An essential part of learning operational rules requires dialogical exchanges with natives. Though it is dangerous to generalize, framing questions in Japan tended to be different than in the United States. For example, in Japan a direct question at an untimely moment can be very destructive, merely by virtue of its directness. Direct questions are to be avoided, reflecting a Japanese value of allowing things to remain implicit rather than calling attention to oneself, making oneself clear and taking responsibility for whatever consequences were incurred. If one asks "Who are you?" in order to obtain an explicit signifier indicating a person's identity as a member of a specific collection or subgroup, one often gets the answer "A Japanese". What appears to be a direct question often becomes an indirect speech act (i.e., those utterances that indicate the appearance of disturbing topics in discourse and therefore require circumlocution.

(Searle, 1975). In fact, the subject is omitted in about 60% of spoken Japanese sentences (Morley, 1985). Gathering answers to personal questions is particularly difficult. Many US visitors remarked about the reticence of their Japanese acquaintances to state what they actually knew.

The tentative, self-effacing, ambiguous use of language is known as *aimai* and this may be attributed to sociological and psychological factors. For one, the historical impact of the autocratic Tokugawa shogunate (1603-1867) where those who practiced verbal discretion definitely had longer life spans. As the old Japanese saying goes, the nail that sticks out gets banged down. A core Japanese value appears to be harmony,

> which they seek to achieve by a subtle process of mutual understanding, almost intuition, rather than by a sharp analysis of conflicting views or by clear-cut decisions, whether made by one-man dictates or majority votes. Decisions, they feel, should not be left up to any one man but should be arrived at by consultations and committee work. Consensus is the goal. (Reischauer, 1981,p.135)

The Japanese itineraries contained more instances of aimai than the American itineraries. American plans were frequently filled with details about desired interactive contacts with individuals whereas the Japanese emphasis seemed to concentrate more on visual experiences with physical sites and organizations versus individuals.

Morley's (1985) analysis of this phenomenon mentions Japanese proverbs testifying to the advantages of silence such as "a defeated warrior does not speak" or *sannen kata-ho*, meaning that a warrior should not allow himself to show emotion more than once in three years. Much is suggested by indirection or vague implication. According to Reischauer (1981), the Japanese have a word for this indirect and/or nonverbal meeting of the mind, *haragei*, whose literal translation is "the art of the belly". Apparently, Americans are still struggling with aimai when attempting to interpret Japan's post World War II constitution, conceptualized under the Marshall Plan in English but open to different interpretations in the operational Japanese version.

Another explanation of aimai can be derived from a remark made by the psychologist Okonogi Keigo is a discussion of *haji*, the Japanese concept of shame. "*Haji* is the characteristic feeling of people who are dependent upon on a group, taking the form of a vague fear of rumour, gossip, of possible ostracism. In the day to day life of people who are sensitive to this kind of shame the essential thing is accordingly not to express oneself unequivocally." Even when translated into English, many American study tour members formed the conclusion that the Japanese approached questions in a deliberate way *not* to be understood by outsiders.

The foreigner or *gaijin* in Japan is treated very politely but always as an outsider--no matter how long he lives in Japan or how deeply he may be involved in its life. Extended interpersonal experiences were avoided by most Japanese that tour members met for the first time. This appeared to be in direct juxtaposition to American preferences. Long time resident and Japanophile Reischauer captured it well when he said:

> If he (foreigner) speaks any Japanese at all, no matter how badly, he is praised for this remarkable accomplishment, as though he were an idiot child who suddenly showed a streak of intelligence. If he is knowledgeable about Japan, he may be asked deferentially his opinions and told he knows more than they do themselves, but his opinions are regarded always as those of an outsider, not an insider. (Reischauer, 1981, p 405)

Is the answer true or false, personal or classificatory when there is a difference in motives between the questioner and the informant and both parties wish to transform their exchang into culturally pertinent material? Umberto Eco (1976, pp. 271-1) uses a game model to describe the process whereby participants in an aesthetic communication accept an underlying set of rules or "system of mutual correlations" and then break or stretch these rules to produce an unanticipated effect. He goes on to state that "the aesthetic text is like a multiple match played by different teams at a time, each of whom follows (or breaks) the rules of their own game" (p. 271). The result, he concludes is " a semiotic design which cunningly gives the impression of nonsemiosis. The aesthetic expression requires a special interpretative leap (p. 38). Both American and Japanese travel itineraries scheduled experiences which seem to predestine being captured in stereotypical tourist rituals of an economic and recreational nature.

Studying the educational system coupled with site visits to schools and interaction with children in each country proved to be an expeditious way of experiencing the game and identifying critical components of semiotic design. All groups scheduled school contacts, though this appeared to be addressed much more concretely by the American tours. Morley (1985) challenges us to observe simple school rituals which offer deep cultural insights given the centralized nature of the Japanese educational system. Consider how children are taught to write letters in Japan. He observed that the nationalized Japanese curriculum informs children that a letter should never begin with oneself but with an inquiry into the health and the general well-being of its recipient. This should be followed by comments about the weather. For each month a set phrase was provided. Only then was it considered proper to broach the real topic of the letter. In our experience as evaluators working with Japanese, this phenomena also occurred with phone and fax inquiries and responses.

For the individual study tour member as well as evaluators, collecting descriptions and judgments of an cultural experience entails similar distinctive literary rules and figurative devices. Empirically dominated strategies such as the objectives based or decision making approaches seem to be disparaged as *rikutsuppoi* or "smacking too much of logic that ignores reality". Coupled with a prevalent Weltanschauung of fatalism and passive expectancy, this explains the frequent use of traditional and predictable events as an approach to cultural awareness by the Japanese itineraries. Perhaps this *rikutsuppoi* accounts in part for the absence of the concept " program evaluator" in Japanese.

Minami Hiroshi posits a national masochism that perceives of acceptance of the negative as spiritual strength. One sees this in numerous Japanese aesthetic productions, notably the work of Yokio Mishima. *Fusoku-shugi*, a fourteenth century theory of insufficiency which identified the most profound aesthetic experience as lying in that which had been withheld and left unsaid, appears very much in the fabric of modern Japan and goes against the grain of Western evaluators.

Consider the rules inherent in communicating with people who assume their language is a means of retention and exclusion (Japanese) versus one of release and expression (American). For example, the courtesy language adopted in Japan for exchanges outside the *uchi* (a social unit related by blood and common locale, but also covering any persons brought together physically and spiritually in a common cause) is circumspect, distrustful, and avoids sensitive topics. American tour members consistently cited this phenomenon. Self-sufficiency exists within one's uchi rather than within oneself. This courtesy language--both oral and written, verbal and nonverbal--is indicated by self-effacement expressions, cautious references, superfluous agreement, solicitude and deference. Interestingly, the rules that controlled behavior in Japan where not always in force when Japanese travelled outside their own country.

The research and evaluation acts of exposure, analysis and critique does not extend beyond one's uchi or organization without resistance. More than a continuous search to sell your skill to the highest bidder, a job in Japan means pride in being part of something big and significant for the rest of your working life. And this group identification is pervasive, extending from schools, especially at the college level, to PTAs, hobby groups, Rotary, new religions, and sightseeing groups (Reischauer, 1981). Given the American values of forthrightness, emotional demonstrations, confrontation, physical strength and individual independence, these Japanese language signs are rejected by US outsiders before they even have time to make them overt. Those gaijin who spoke fluent Japanese went through dissociation from their own culture, even a sort of hostility on the road to their fluency and recounted tales of how Japanese had refused to hear them speak Japanese and tried to answer in English rather than recognize a *gaijin*'s ability to play the game. A solution for even fluent speakers is the wide use of go-betweens and established network of contacts. The go-between scouts out the views of both sides, finds ways around problems and if the emotions become too vivid, end the exchange without the danger of a direct (and attention drawing) confrontation. On the otherhand, Americans seemed all to eager (or relieved?) to discover Japanese typically spoke much better English and over time came to expect this facility.

What emerges is the critical importance of fluency in verbal *and* nonverbal language for both insiders and outsiders. Even if one masters verbal Japanese replete with its numerous incomplete sentences, there are so many ambiguous pre-speech mood orienting sounds required for one's own productions or encouraging others expression (*aizuchi*), but little spontaneous and explicit body language which seems to be the reverse of Americans. The closest analogy we can

think of in the American scene is the interrogatory dissembling of the TV detective, Columbo. Accompanying "appropriate" verbal and nonverbal exchanges are its accoutrements such as context appropriate dress or uniforms in keeping with the occasion (*basho-gara wakimaezu*) and the exchange of meishi (*business cards*) at even the most apparently casual encounter. Meishi helped fill the semiotic gaps by categorizing the individual by name, organization, rank and function. In the US we have no such mechanism, the absence of which suggests another cultural sign.

Major learnings and emerging polemics

Evaluation is ever present and will continue to operate, both overtly and covertly. The evaluation is only as good as the shared Umwelt that the evaluator is able to coalesce and make overt among all significant players. Approaching this 'willed' reality or new Umwelt can be facilitated by Polanyi's (1962, 1966, 1975, 1978) constructs of collecting multiple perceptions in order to approach a shared reality, trying to make tacit perceptions overt rather than being unconsciously controlled by these perceptions, and extended dwelling in an experience to experience more than the surface phenomena. In essence, out of these diverse perceptions we can dentifyi critical polemics as signposts which expand but do not fix our capability to describe and judge. By presenting thirteen polemics emerging from our bi-national evaluation experiences as signs, we hope to stimulate discussion of how semiotics informs evaluation inquiry. Table 2 summarizes these polemics.

---

insert Table 2 about here

How can balanced efforts to "improve" educational programs be achieved in changing political-economic climates when these activities are supported by diverse intra-national and international purposes? One of the major challenges of this evaluation quickly became the management of nine geographically diverse sets of quantitative and qualitative data in the United States and two in Japan. How does an external evaluation communicate this morass of local, regional and national programming to a bi-national board of trustees to facilitate *unified* futures planning--in essence, to negotiate a shared Umwelt? The use of polemics proved very helpful here.

Workable notions of "core cultural curricula" and "global education" have been only partially realized world-wide. The road to global concepts such as bi-culturalism within curricular contexts is sequential. Initial emphasis must be given to localization. Ultimately, this examination has been targeted on the affordability of empiricism and "proof" versus "improvement" paradigms in more localized and regional settings. It appears that the program leadership must constantly balance an egalitarian versus an elitist polemic. In the extent of our evaluation experiences,

enhanced cultural awareness and curriculum infusion must replace established systems of material and symbolic exchanges between political and educational elites, communities, parents and direct service providers--the teachers. The more the country is in flux economically, the more the question of value imposition resulting from reform movements looms. It is definitely a challenge for the foundation to be both eclectic and capable of listening to diverse values, ecological in light of limited resources and personnel and identifying common bi-national values so that a synthetic base of "program improvement" can be perceived as helpful.

What kind of evaluation paradigm (or myth according to semioticians) is valuable? Value must be considered in some kind of culturally sensitive semiotic context, making cross-national applications difficult to achieve. For transferability of evaluation, trust must be present. This is a prospect difficult to achieve in one's own culture, let alone attempting to cross East-West boundaries. Much of the actual use of evaluation in the US would seem to have limited applicability in Japan though perception-based models would appear to be the most consonant. "Levels of consciousness" and ethnocentricity appear to influence significantly the concepts, applications, timeline, and ultimate usefulness of evaluation. Even in the US, the struggle to orient program evaluation to an improvement-directed stance is constant; many evaluation audiences have strongly reinforced perceptions of evaluations being exclusively directed toward "proof" of programs. Teaching the evaluated that perceptions determine their reality is a challenge that must be confronted early (as well as determining what weight will be accorded to representative perceptions of all the major stakeholders).

Who verifies the procedures of translation and evaluation? And how does the evaluator discern accurate information from misinformation? Avoiding manipulation and achieving holism demands risk-taking on the part of the evaluated and the evaluators alike. Given that we are dealing with two cultures with distinctly different core value systems, the process of paradigm creation cannot rely solely in the hands of one evaluator. Multiple perceptions of diverse audiences in both cultures must be obtained in order to more accurately regulate the interaction of assumptions and representations for recognition. If we assume Benzon & Hays' (1987) contention that physiognomic and visually holistic representations tend to be the preferred mode in Eastern cultures and Western perceptions tend to display propositional and verbal representations, 360 degrees of perceptual descriptions must be collected in an effort to make the implicit explicit. In fact, each distinct constituency represented in the evaluation risks losing some of their unique commodity in the process of negotiating a greater shared reality.

The discrepancies between pre-and post-trip perceptions leads you to consciously seek the myth or paradigm which is built upon verbal *and* iconic parts. You then search for the part where the discrepancies are greatest, that will tell you where your awareness has changed. But what if you are like Ann Tyler's *Accidental Tourist*, exposed to other cultures but not seeking

discrepancies (in fact, assiduously avoiding them) because you have been captured in your own cultural cacoon? Or what happens when the traveller and the native informant power consciously hold contrary positions? Potentially, what is at risk is the transformation of the knowledge base of the traveller versus the transformation of the native. A discrepancy analysis remains incomplete without negotiation of perceptions of diverse audiences and stakeholders. Instruction in cultural metaphors can lead to recognition and over time to cultural awareness and fluent reading of signs. If we merely consider the dual meaning of the concept "alien", we may conclude that practice in reading signs in other cultures helps develop a conscious reading in one's own culture.

Summary

Even within educational program evaluation, the major paradigms fall short of a theory that meets the semiotic criteria set forth in this paper. When one examines the works of evaluators such as Stuffelbeam one finds the necessary components of a paradigm. While their models are mechanically total, they exclude the negotiation of values when major questions of focus are being determined as well as fail to predefine or make overt the values to be applied in making judgments. Michael Scriven, a philosopher as well as an evaluation theorist, deals with values but his external approach assumes an elitist orientation. Scriven suggests that values must be brought into the schema from outside sources for judgments to be made, but does not set up the necessary dynamics of negotiating major questions to focus the study. Robert Stake is concerned about judgments and values and does include the "outsider" evaluator's relative and absolute judgements while avoiding the pitfalls of an absolutist position regarding a program's goals.

A semiotic paradigm is clarified by the conscious inclusion of pr.emics in order to explicate program myths. In essence, we propose approaching evaluation as both a *native* and an *alien*, an insider and a disciplined outsider. What is needed is a paradigm which negotiates major questions openly with all participants; gathers quantitative and qualitative sources of evidence; and applies standards or value criteria which are neither elitist from the point of view of the evaluator, nor captured by the tacit agendas of the stakeholders.

Table 1. **The Perception-Based Evaluation: a Semiotic Paradigm**

Box 1	Box 2	Box 3
What major questions should be asked in the evaluation?	**What sources of evidence can be found?**	**What quality/value components will be applied?**
Questions having a holistic emphasis wherein all syntheses are repeatedly tested against experiential perceptions Questions which are helpful through a dynamic dialogue, negotiation, and pedagogical action or make visible the values being synthesized by both evaluator and evaluatee Questions which expand the range of major questions asked by recognizing distortions of positivist and reductionist biases Questions which emphasize the vulnerability of the evaluator to evaluate in an effort to avoid oppression of evaluatee Questions which facilitate the planning of program futures	Multiple methodologies are used in an effort to survey the "whole" of reality: e.g., descriptions, observations, subjective impressions, scientific method, interviews, perceptions, & artifacts Should be those sources which are closest in agreement with multiple perceivers Both "hard" and "soft" data are systematically used to obtain an accurate picture of reality	Judgments and values grounded in perceptions and are overt Openness to seeking multiple visions of reality and the totality of experience Orientation to the future, extendability, and transferability of levels of consciousness Discussion, mutual responsibility, mutual criticism, and negotiation Respect for core values

Table 2. Polemics of an Evaluation Semiotic

1. holistic dwelling	( )	( )	( )	( )	( ) one-shot reviews
2. paternal evaluator	( )	( )	( )	( )	( ) negotiating evaluator
3. client helpful	( )	( )	( )	( )	( ) field helpful
4. obtrusive	( )	( )	( )	( )	( ) intrusive
5. quantitative	( )	( )	( )	( )	( ) qualitative
6. descriptions	( )	( )	( )	( )	( ) judgments
7. focus	( )	( )	( )	( )	( ) power
8. positive orientation	( )	( )	( )	( )	( ) negative orientation
9. tacit	( )	( )	( )	( )	( ) overt
10. independent	( )	( )	( )	( )	( ) negotiated
11. proactive	( )	( )	( )	( )	( ) reactive
12. initiated	( )	( )	( )	( )	( ) imposed
13. accountable	( )	( )	( )	( )	( ) subordinate

REFERENCES

AAAS Commission on Science Education. An Evaluation Model and Its Application. Washington, D.C.: American Association for the Advancement of Science. Publication 65/9, 1965.

Alkin, M.C. (1968). Towards an Evaluation Model: A System Approach. CSE Working Paper, No. 4. Los Angeles, CA: Center for the Study of Evaluation.

Allo, G. (1961). Research on Values at the Crossroads of Modern Civilizations. Beirut: Institute de Recherce et de Formation en Vue du Developpement (IFRED)

Ary, D., Jacobs, L., & Razavieh, A. (1972). Introduction to Research in Education. New York, NY: Holt, Rinehart and Winston.

Baron, N.S. (1987). When seeing's not believing. The American Journal of Semiotics, 5(3 & 4), 321-339.

Benzon, W.L. & Hays, D.G. (1987). Metaphor, recognition, and neural process. The American Journal of Semiotics, 5(1), 59-80.

Bergson, H. (1910). Essai sur les donnes immediates de la conscience. In F.L. Pogson (Ed.), Time and Free Will. New York, NY: The Macmillan Co.

Campbell, D.T. (1984) Foreward. In R.K Yin, Case study research. Beverly Hills, CA: Sage Publications.

Cronbach, L.J. (1963). Course improvement through evaluation. Teacher College Record, 64, 672-683.

Cunningham, D.J. (1987). Outline of an education semiotic. American Journal of Semiotics, 5(2), 201-216.

Dillon, G. (1981). Constructing Texts: Elements of a Theory of Composition and Style. Bloomington, IN: Indiana University Press.

Driscoll, M.P. & Dempsey, J.V. (1988, April). An authoring template for computer-based coordinate concept instruction. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Driscoll, M.P. & Flynn, J. (1989) Conducting semiotic educational research: a semiotic analysis of an educational study. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Earle, W. (1955). Objectivity: Essay in Phenomenological Ontonogy. Chicago, IL: Quandrangle Paperbacks.

Eco, U. (1976). A Theory of Semiotics. Bloomington, IN: Indiana University Press.

Freire, P. (1970 a.). Pedagogy of the Oppressed. New York, NY: Herder and Herder.

Freire, P. (1970 b.). Cultural Action for Freedom. Harvard Educational Review, Monograph No. 1.

From, F. (1971). Perception of Other People. New York, NY: Columbia University Press.

Garfinkel, H., & Sacks, H. (1970). On the formal structures of practical actions. In J. McKinney and E. Tiryakion (Eds.) Theoretical Sociology. New York, NY: Appleton-Century Crofts.

Geertz, C. (1973). The Interpretation of Cultures. New York, NY: Basic Books.

Gordon, T.J. (1969). The feedback between technology and values. In K. Baier and N. Rescher (Eds.). Values and the Future. New York, NY: The Free Press.

Goulet, D.A. (1968). Development for what? Comparative Political Studies, 1, 295-312.

Goulet, D.A. (1971). The study of values. Harvard Educational Review, 41(2), 205-227.

Graubard, A. (1972). Free the children. New York, NY: Pantheon Books.

Guba, E. G. (1969). The failure of educational evaluation. Educational Technology, 9(5), 29-38.

Gurwitsch, A. (1966). Studies in Phenomenology and Psychology. Evanston, IL: Northwestern University Press, 428-432.

Houser, N. (1987). Toward a Peircean semitic theory of learning. The American Journal of Semiotics, 5(2), 251-274.

Husserl, E. (1964). The Phenomenology of Internal Time--Consciousness. The Hague: Nijhoff.

Huxley, A. (1963). The Doors of Perception. New York, NY: Harper and Row.

Jules-Rosette, B. Expression and aesthetics in science and art: ethnography as discursive sabotage. The American Journal of Semiotics, 6(1), 37-55.

Kuhn, H. (1962). The Structure of Scientific Revolutions. Chicago, IL: The University of Chicago Press.

Kunkel, R.C. & Tucker, S.A. (1983). Critical evaluation skills at a time of retrenchment. Paper presented at the annual meeting of the Association for Supervision and Curriculum Development, Houston, TX.

Landgrebe, L. (1940). The world as a phenomenological problem. Philosophy and Phenomenological Research, I.

Levi-Strauss, C. (1969). The Raw and the Cooked. New York, NY: Harper and Row.

Levi-Strauss, C. (1976). Structural Anthropology. Vol II. New York, NY: Harper and Row.

Levi-Strauss, C. (1979). Myth and Meaning. New York, NY: Schocken Books.

Merleau-Ponty, M. 1962). Phenomenology of Perception. London, UK: Rutledge and Kegan Paul.

Merleau-Ponty, M. (1964). The Primacy of Perception: and Other Essays. Evanston, IL: Northwestern University Press.

Merleau-Ponty, M. (1965). The Structure of Behavior. London, UK: Methuen.

Morley, J.D. (1986). Pictures from the Water Trade: Adventures of a Westerner in Japan. NewYork, NY: Perennial Library.

Pace, R. C. (1968, December). Evaluation perspectives: 1968. Los Angeles, CA: UCLA Center for the Study of Evaluation. (ERIC ED 037828).

Pharies, D.A. (1985). Charles S. Peirce and the linguistic sign. Amsterdam/Philadelphia, PA: John Benjamin Publication Co.

Polanyi, M. (1962) Personal Knowledge. Chicago, IL: University of Chicago Press.

Polanyi, M. (1966). The Tacit Dimension. London. UK: Pergamon Press.

Polanyi, M. (1975). Meaning. Chicago, IL: University of Chicago Press.

Polanyi, M. (1978). Phenomenology of Perception. Tr.C.Smith. London, UK: Routledge and Kegan Paul.

Porsecanski, T. (1989). Ideologies and development: a report from South America. The American Journal of Semiotics, 6(2/3), 291-298.

Provus, M. (1969). Evaluation on ongoing programs in the public school system. In R.W. Tyler. (Ed.) Educational Evaluation: New Roles, New Means. Chicago, IL: N.S.S.E.

Provus, M. (1971). Discrepancy Evaluation. Berkeley, CA: McCutchan Publishing Company.

Reasby, H. (1973). Application of multiple evaluation strategies to comprehensive and innovative middle school programs. Unpublished Doctoral Dissertation, University of Washington.

Reischauer, E.O. (1981). The Japanese. Cambridge, MA: The Belknap Press of Harvard University Press.

Ricoeur, P. (1985). Time and Narrative. Tr. K. McLaughlin & D. Pellauer. Chicago, IL: University of Chicago Press.

Riffaterre, M. (1987). Hypersigns. The American Journal of Semiotics, 5(1), 1-12.

Sacks, H. (1972). On the analyzability of stories by children. In J.J. Gumperz & D. Hymes (Eds.), Directions in Sociolinguistics: The Ethnography of Communication. New York, NY: Holt, Rinehart and Winston, 325-345.

Schutz, A. (1967). The Phenomenology of the Social World. Evanston, IL: Northwestern University Press.

Scriven, M. (1967). The methodology of evaluation. In B.O. Smith. (Ed.), AERA Monograph Series on Curriculum Evaluation. Chicago, IL: Rand McNally.

Scriven, M. (1972). Prose and cons about goal-free evaluation. Evaluation Comment, 3(4), 1-8.

Scriven, M. (1978). Goal-free evaluation in practice. Paper presented at the fourth annual meeting of the American Educational Research Association, Toronto, Ontario.

Searle, J. (1975). Indirect speech acts. In P. Cole & J.L. Morgan (eds.) Syntax and Semantics, Vol.3: Speech Acts. New York, NY: Academic Press, 59-82.

Sebeok, T.A. (1986). A semiotic perspective in the sciences: Steps toward a new paradigm. In I think I am a verb. New York, NY: Plenum Press.

Stake, R. E. (1967). The Countenance of educational evaluation. Teachers College Record, 68(7), 523-540.

Stake, R.E. (1970). Objectives, priorities, and other judgment data. Review of Educational Research, 40(2), 181-212.

Stake, R.E. (1972). Focus or portrayal. Evaluation and Measurement Newsletter, 14, 1-2.

Stake, R.E. (1973). Measuring what learners learn. In E. R. House (Ed.) School Evaluation. The Politics and Process. Berkeley, CA: McCuthan Publishing Co.

Stake, R.E. (1975). Program evaluation, particularly responsive evaluation (Occasional Paper, No.5). Kalamazoo, MI: Western Michigan University Evaluation Center.

Stake, R.E. (1982). A peer response: A review of program evaluation in education. When? How? To what ends? In E.R. House, S. Mathison, J.A. Pearsol, & H. Preskill (Eds.), Evaluation studies review annual (vol 7). Beverly Hills, CA: Sage.

Stufflebeam, D.L. (1968). Toward a science of educational evaluation. Educational Technology, 8(14), 5-12.

Stufflebeam, D.L. (1983). The CIPP model for program evaluation. In G.F. Madaus, M. Scriven, & D.L. Stufflebeam (Eds.), Evaluation Models: Viewpoints on Educational and Human Services Evaluation. Boston, MA: Kluwer-Nijhoff.

Stufflebeam, D.L., Foley, W.J., Gephard, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., & Provus, M. M. (1971). Educational Evaluation and Decision-making. Itaska, IL: F. E. Peacock Publishers, Inc.

Taba, H., & Sawin, E. (1962). A proposed model in evaluation. Educational Leadership, 1962, 57-71.

Tyler, R.W. (1942). General statement on evaluation. Journal of Educational Research, 30, 492-501.

Weckworth, V.E. (1969). A conceptual model versus the real world of health care service delivery. In Hopkins, C.E. (Ed.) Outcomes Conference: Methodology of identifying, measuring and evaluating outcomes of Health Service Programs, Systems and Subsystems. Rockville, MD: Office of Scientific and Technical Information, NCH SRD, 39-62.